Vragen? Opmerkingen? Tips?
Mail ze naar
brieven@clickx.be.

DOE HET ZELF

Brief van de week

De Clickx-redactie wordt elke dag overstelpt met vragen van lezers. Sommige problemen zijn te specifiek om in het magazine te behandelen, maar andere vragen zijn dan weer zo interessant dat ze meer verdienen dan een kort antwoordje. Daarom selecteren we voor elke Clickx Magazine een vraag van een lezer, die we dan uitwerken in een complete workshop. De vraag vind je op deze pagina, de workshop staat op de volgende twee pagina's. Veel plezier!

GEVANGEN IN GOOGLE



Op mijn website staat een naam van iemand die zijn naam daar liever niet ziet verschijnen. Ik heb zijn naam dus verwijderd, maar als ik zijn naam ingeef via Google, verschijnt er toch nog een link naar mijn website. De naam is immers nog te vinden in het cachegeheugen van de zoekmotor. We zijn nu meer dan één maand verder en nog steeds staat die pagina in de Google-cache. Hoe los ik dat op?

KOEN TAILLIEU



Een opmerkelijke vraag, Koen, want de meeste mensen mailen ons om te vragen hoe hun site wél kan opgepikt worden door Google en hoe ze zo hoog mogelijk scoren in de zoekresultaten. Op de volgende pagina's geven we je wat tips over welke aanpassingen je zoal moet doorvoeren om er voor te zorgen dat de Googlebot je webstekje (of een deel ervan) voorbijgaat.





GEVANGEN IN GOOGLE

WAT DOEN WE?
- INFORMATIE UIT DE GOOGLE-ZOEKRESULTATEN VERWIJDEREN EN JE SITE (DEELS) AFSCHERMEN VOOR DE GOOGLEBOT

WAARMEE?
- INTERNET, KLADBLOK

DUUR?
- 30 MINUTEN

MOEILIJKHEID?



Voor alle duidelijkheid: je kan alleen maar informatie uit Google verwijderen als die gegevens op een site staan die je zélf beheert. Als je naam ergens op een site opduikt waar je zelf geen bestanden aan kan wijzigen, en je wil die verwijderd zien, hangt alles af van de *goodwill* van de webmaster in kwestie. En dan nog. Want zelfs als je een eigen site hebt en netjes alle stappen volgt die hieronder beschreven staan, heb je nog geen garantie dat de gewraakte gegevens niet ergens opduiken in de resultaten van de zoekmachine. Googles wegen zijn ondoorgrondelijk... Op hoop van zegen dus!

STAP 1 / VERWIJDER ZELF DE GEGEVENS

Het lijkt misschien een overbodige stap, maar voor je allerlei ingrepen begint uit te halen om informatie uit de zoekresultaten van Google te weren, verwijder je best ook eerst zelf de gewraakte gegevens van je site. Je weet zelf het best hoe je dat doet, want geen enkel webbouwgereedschap is hetzelfde. Pas dus alle gegevens aan, en na verloop van tijd (meestal enkele weken) zal ook de Googlebot dit opgemerkt hebben en zijn gegevens aanpassen. Maar daarmee ben je nog niet uit de cache van Google verwijderd. Google bewaart – net als heel wat andere zoekmachines – alle pagina's die het indexeert namelijk ook nog op zijn eigen servers.

STAP 2 / FORMULIERTJE INVULLEN

Als de live informatie van je site is verwijderd, kan je bij Google een formuliertje invullen en hen vriendelijk vragen om bepaalde gegevens uit hun cachegeheugen te verwijderen. Surf naar http://google.com/support/bin/request.py?contact_type=remove_info (zie afbeelding 2). Geef je naam en je e-mailadres in, samen met de url van de pagina die opduikt in de Google-resultaten. Let wel: dit moet het exácte webadres zijn. Om zeker te zijn, geef je op www.google.be dus nog best even de zoekterm in waarmee je op die specifieke pagina terechtkomt. Klik op het adres van de website en kopieer het uit je adresbalk met **Ctrl+C**. Plak het nu in het veld **URL** met de toetsencombinatie **Ctrl+V**. Geef een onderwerp in in het veld **Subject**, leg het probleem gedetailleerd uit bij **Message** (doe dit wel in het Engels) en klik op **Submit**. Als alles duidelijk is, moet je geen bericht van Google verwachten en wordt je vraag binnen de kortste keren behandeld. Maar slimmeriken die denken dat ze hun concurrenten op deze manier een loer kunnen draaien, moeten we teleurstellen. Er zijn namelijk nog heel wat andere factoren die meespelen, maar die houdt Google angstvallig geheim.

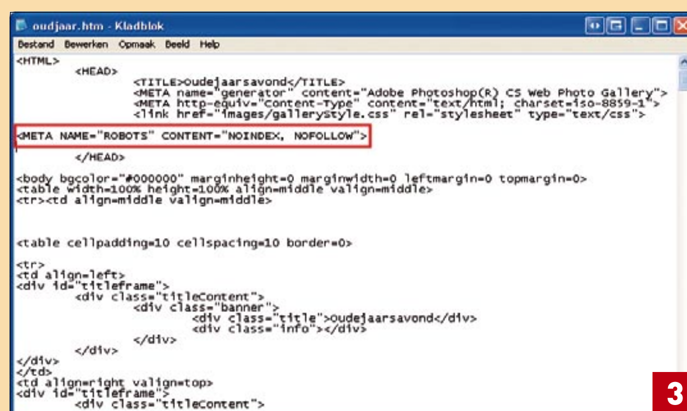
Beleefd als we zijn, vragen we vriendelijk aan Google om onze gegevens uit hun cachegeheugen te verwijderen.

STAP 3 / INDIVIDUELE PAGINA'S AFSCHERMEN

Het probleem van lezer Koen Taillieu zou met deze 2 stappen opgelost moeten zijn – al kan het wel even duren vooraleer de wijzigingen echt zichtbaar zijn. In de rest van deze workshop gaan we proactief te werk en vertellen we hoe je kan voorkomen dat zo'n probleem zich stelt. We beginnen bij de kleinst mogelijke stap: een individuele pagina verstoppen voor de Googlebot en andere zoekrobotten. Hiervoor moet je een aanpassing maken in de broncode van de pagina in kwestie. Klik met je rechtermuisknop op het html-bestand en kies **Openen met**. Vervolgens selecteer je **Kladblok**, en de broncode wordt geopend. Ga nu op zoek naar de tags **<head>** en **</head>**. Ergens tussen deze twee tags zet je nu het volgende stukje code (zie afbeelding 3):

```
<META NAME="ROBOTS" CONTENT="NOINDEX, NOFOLLOW">
```

Hiermee geef je te kennen dat alle zoekrobotten de inhoud van die pagina niet mogen indexeren (**NOINDEX**) en dat ze ook de links op die pagina niet mogen volgen (**NOFOLLOW**). Wil je alleen de Googlebot buitenhouden, dan vervang je **ROBOTS** door **GOOGLEBOT**. Upload het gewijzigde html-bestand naar je webserver en overschrijf het originele bestand.



Als je een bepaalde metatag aan een pagina toevoegt, knijpen zoekrobots voortaan een oogje dicht als ze langskomen.

STAP 4 / NIET IN DE CACHE!

Zoals gezegd houden zoekmachines vaak een kopie van geïndexeerde webpagina's bij op hun eigen servers. Op zich een zeer handige service voor surfers, want zelfs nadat een site is opgehouden te bestaan, kan je de informatie van die site nog altijd terug opvragen (zie afbeelding 4). Als je wil dat je pagina in de zoekresultaten terechtkomt, maar niet dat die lokale kopie gemaakt wordt, open je opnieuw de broncode van

je pagina (zie ook stap 3) en plaats je het volgende stukje code tussen de head-tags:

```
<META NAME="ROBOTS" CONTENT="NOARCHIVE">
```

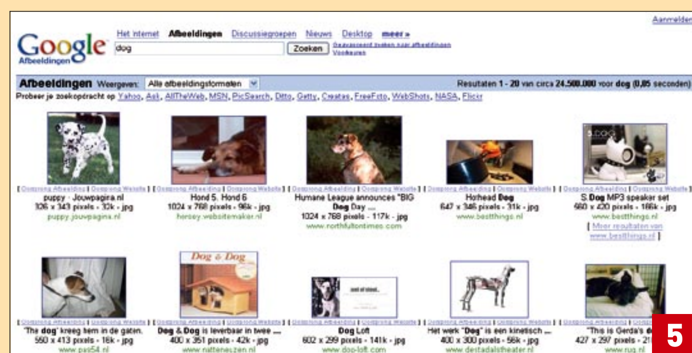
Zoals je al kan vermoeden, kan je ook hier **ROBOTS** vervangen door **GOOGLEBOT** als je wil dat alleen Google geen lokale kopie maakt. Vergeet het bestand niet te uploaden als je klaar bent.



Zelfs al is een pagina van je eigen webserver verwijderd, dan nog kan Google die uit zijn cachegeheugen tevoorschijn toveren.

STAP 5 / ALLE AFBEELDINGEN BESCHERMEN

Google Afbeeldingen is voor veel surfers een uitstekende bron als ze beeldjes van om het even wat nodig hebben (zie afbeelding 5). Maar ook hier kunnen er wel eens problemen met auteursrechten opduiken. Webmasters die niet zozeer teksten maar wel foto's op hun site aanbieden, vinden het immers niet leuk dat hun beeldjes zomaar op andere sites opduiken, zonder bronvermelding. En snoodaards die aan *hotlinking* doen, krijgen vrij spel met de afbeeldingen die ze via de plaatjeszoeker van Google vinden. Je hebt er dus alle belang bij om ook je foto's te beschermen. Om dit te bekomen, moet je een bestand aanmaken dat



Als je niet wil dat jouw zelfgemaakte afbeeldingen in Google Images belanden, kan één eenvoudig bestandje dat veranderen: robots.txt.

luistert naar de naam robots.txt. Vooraleer zoekrobotten een site uitspitten, gaan ze op zoek naar de 'richtlijnen' die in dit bestand staan. De werkwijze is relatief eenvoudig. Maak een nieuw txt-bestand aan, bijvoorbeeld door met je rechtermuisknop op je bureaublad te klikken en in het snelmenu **Nieuw**, **Tekstdocument** te kiezen. Geef het bestand meteen de juiste naam (**robots.txt**) en dubbelklik erop om het te openen. In het maagdelijk witte tekstbestand tik je nu:

```
User-Agent: Googlebot-Image
Disallow: /
```

Hiermee worden alle beelden uit de index van Google Afbeeldingen geweerd. Bewaar het bestand en upload het naar de server waarop je website staat. Let wel: robots.txt moet in de rootfolder van je site staan. Doe zelf de test en surf naar **www.jouwsite.be/robots.txt**. Als je daar de inhoud ziet staan, staat het op de juiste plaats.

STAP 6 / SPECIFIEKE AFBEELDINGEN BESCHERMEN

Het zou ook kunnen dat je niet álle beeldjes op je site wil beschermen, maar bijvoorbeeld alleen je vakantiefoto's. Stel dat je die op je website in de submap /vakantie2007/ hebt gezet. Ook dat kan je dan via het robots.txt-bestandje te kennen geven aan de Googlebot. In het voorbeeld van hierboven geef je dan de opdracht:

```
User-Agent: Googlebot-Image
Disallow: /vakantie2007/
```

Vergeet zeker niet de slash (/) achteraan de mapnaam; anders denkt de server dat het om een enkel bestand gaat – wat trouwens ook mogelijk is: zet bij **Disallow** dan gewoon het pad van het ene bestand, inclusief extensie (bijvoorbeeld **/vakantie2007/foto1.jpg**). Stuur na het uploaden van dit bestand een mailtje met de verandering en de locatie van de foto's naar **googlebot@google.com**, en Google verwijdert ze binnen de 48 uur uit zijn index.

STAP 7 / JE HELE SITE BLOKKEREN

En last but not least: als je wil dat Google en alle andere zoekmachines je site gewoon voorbijgaan, en absoluut niets ervan indexeren, moet je dat ook aangeven in robots.txt:

```
User-Agent: *
Disallow: /
```

Van zodra dit bestand in de rootmap van je server belandt, blijft je site volledig onder de radar en vind je er binnen de kortste keren niets meer van terug in zoekmachines. We snappen niet goed waarom iemand dát zou willen, maar het kan dus wel degelijk. ♦



En als je dat écht wil, kan je er voor zorgen dat je helemaal niet voorkomt in de zoekresultaten van Google.

VAKTAAL A - M N - Z

HOTLINKING: Hotlinking, inline linking of leeching is een techniek om bandbreedte te stelen. De hotlinker zet bijvoorbeeld een afbeelding op zijn site, maar host die niet zelf en verwijst (zonder medeweten en toelating van de eigenaar) naar de site waarop het beeld aanvankelijk stond. De surfer ziet geen verschil, maar het dataverkeer wordt wel 'aangerekend' aan diegene op wiens server het beeld staat.